Selecting Funny Videos on YouTube via Google Developer API

Bochao Jia
Department of Biostatistics, University of Florida

1   Method

The problem is to select the top videos that makes users laugh. Therefore, the most intuitive idea is to retrieve the information of video and users' feedback to judge the quality of videos. To achieve this idea, we applied the Google Developer API on python to retrieve the information on number of likes, dislikes views and the video title given a certain website. However, Youtube has billions of videos and it is not realistic to include all of them into our dataset, therefore, the first step will be search the videos by the key words. Since the problem is to find the videos that makes users laugh, the key words can be "funny Vines", "laugh", etc. In this step, our code can return about 30 websites given a key word, hence, we can try different key words and include all returned websites. The next step is to retrieve the information of video and the feedback of users. Our code can obtain the number of likes, dislikes views and the video title given the website in the first step. Finally, in order to get the list of the top 10 funnest videos and judge the quality of them, we calculate the ratio of likes/dislikes, likes/views, (likes+dislikes)/views which can be used to evaluate the feedback of the users.

2   Results

After we calculated the values of ratios for each video, we need to select one of them to be sorted by for our datasets. In order to choose it, we gives the log-log plot of each pair of variables including number of likes, dislikes, views and ratios as showed in Figure 1. It shows that all pairs of variables are linearly correlated under the log-transformation except the ratio of likes and dislikes. Therefore, the ratio of likes and dislikes can be a unique criterion which used to judge the quality of funny videos. In conclusion, the top 10 videos that make users laugh should be the ones with 10 largest values for the ratio of likes and dislikes, which can be output in an attached csv files.
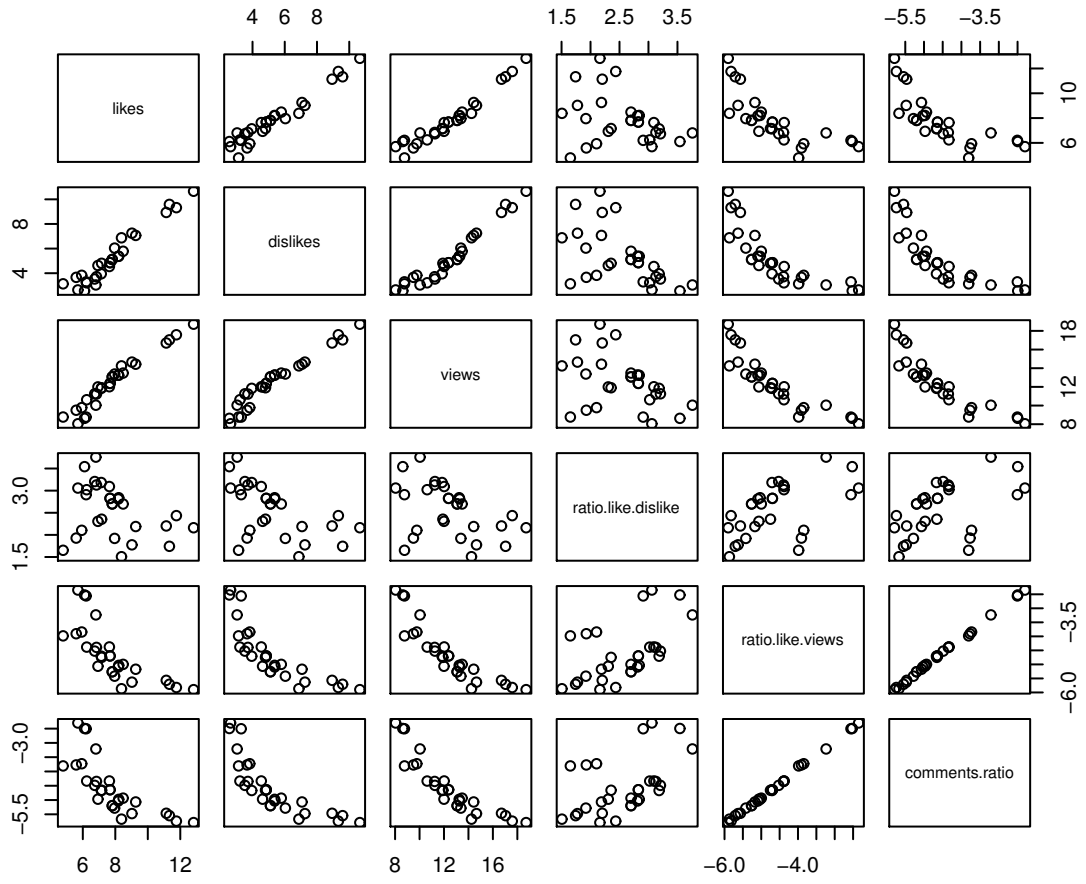
Figure 1: The log-log plot of each pair of variables including number of likes, dislikes, views and ratios.